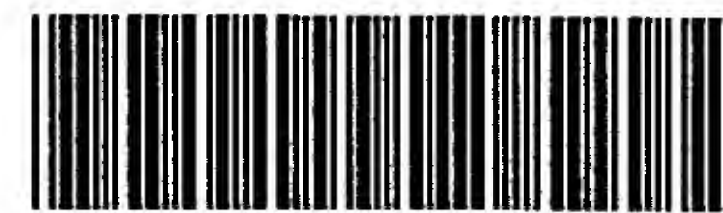# RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: _09/684,305_

Source: _1Fw/6_

Date Processed by STIC: _2/6/05_

ENTERED

IFW16

```
 3 <110> APPLICANT: Kaiser, Michael W.
 4        Lyamichev, Victor I.
 5        Lyamichev, Natasha
 7 <120> TITLE OF INVENTION: Improved Cleavage Agents
 9 <130> FILE REFERENCE: FORS-04447
11 <140> CURRENT APPLICATION NUMBER: 09/684,305
12 <141> CURRENT FILING DATE: 2000-10-06
14 <160> NUMBER OF SEQ ID NOS: 188
16 <170> SOFTWARE: PatentIn Ver. 2.0
18 <210> SEQ ID NO: 1
19 <211> LENGTH: 2506
20 <212> TYPE: DNA
21 <213> ORGANISM: Thermus aquaticus
23 <400> SEQUENCE: 1
24 atgagggggga tgctgcccct ctttgagccc aagggccggg tcctcctggt ggacggccac 60
26 cacctggcct accgcacctt ccacgccctg aagggcctca ccaccagccg ggggggagccg 120
28 gtgcaggcgg tctacggctt cgccaagagc ctcctcaagg ccctcaagga ggacggggac 180
30 gcggtgatcg tggtctttga cgccaaggcc ccctccttcc gccacgaggc ctacggggggg 240
32 tacaaggcgg gccgggcccc cacgccggag gactttcccc ggcaactcgc cctcatcaag 300
34 gagctggtgg acctcctggg gctggcgcgc ctcgaggtcc cgggctacga ggcggacgac 360
36 gtcctggcca gcctggccaa gaaggcggaa aaggagggct acgaggtccg catcctcacc 420
38 gccgacaaag acctttacca gctcctttcc gaccgcatcc acgtcctcca ccccgagggg 480
40 tacctcatca ccccggcctg gctttgggaa aagtacggcc tgaggcccga ccagtgggcc 540
42 gactaccggg ccctgaccgg ggacgagtcc gacaaccttc ccggggtcaa gggcatcggg 600
44 gagaagacgg cgaggaagct tctggaggag tgggggagcc tggaagccct cctcaagaac 660
46 ctggaccggc tgaagcccgc catccgggag aagatcctgg cccacatgga cgatctgaag 720
48 ctctcctggg acctggccaa ggtgcgcacc gacctgcccc tggaggtgga cttcgccaaa 780
50 aggcgggagc ccgaccggga gaggcttagg gcctttctgg agaggcttga gtttggcagc 840
52 ctcctccacg agttcggcct tctggaaagc cccaaggccc tggaggaggc cccctggccc 900
54 ccgccggaag gggccttcgt gggctttgtg ctttcccgca aggagcccat gtgggccgat 960
56 cttctggccc tggccgccgc cagggggggc cgggtccacc gggcccccga gccttataaa 1020
58 gccctcaggg acctgaagga ggcgcggggg cttctcgcca agaacctgag cgttctggcc 1080
60 ctgagggaag gccttggcct cccgcccggc gacgacccca tgctcctcgc ctacctcctg 1140
62 gaccctcca acaccacccc cgagggggtg gcccggcgct acggcggggga gtggacggag 1200
64 gaggcgggggg agcgggccgc cctttccgag aggctcttcg ccaacctgtg ggggaggctt 1260
66 gagggggagg agaggctcct ttggctttac cgggaggtgg agaggcccct ttccgctgtc 1320
68 ctggcccaca tggaggccac ggggggtgcgc ctggacgtgg cctatctcag ggccttgtcc 1380
70 ctggaggtgg ccgaggagat cgcccgcctc gaggccgagg tcttccgcct ggccggccac 1440
72 cccttcaacc tcaactcccg ggaccagctg gaaagggtcc tctttgacga gctagggctt 1500
74 cccgccatcg gcaagacgga gaagaccggc aagcgctcca ccagcgccgc cgtcctggag 1560
76 gccctccgcg aggcccaccc catcgtggag aagatcctgc agtaccggga gctcaccaag 1620
78 ctgaagagca cctacattga ccccttgccg gacctcatcc accccaggac gggccgcctc 1680
80 cacacccgct tcaaccagac ggccacggcc acgggcaggc taagtagctc cgatcccaac 1740
```

RAW SEQUENCE LISTING   DATE: 02/06/2005
PATENT APPLICATION: US/09/684,305   TIME: 12:32:05

Input Set : **A:\04447.app**
Output Set: **N:\CRF4\02042005\I684305.raw**

```
82  ctccagaaca tccccgtccg caccccgctt gggcagagga tccgccgggc cttcatcgcc 1800
84  gaggagggt ggctattggt ggccctggac tatagccaga tagagctcag ggtgctggcc 1860
86  cacctctccg gcgacgagaa cctgatccgg gtcttccagg aggggcggga catccacacg 1920
88  gagaccgcca gctggatgtt cggcgtcccc cgggaggccg tggaccccct gatgcgccgg 1980
90  gcggccaaga ccatcaactt cggggtcctc tacggcatgt cggcccaccg cctctcccag 2040
92  gagctagcca tcccttacga ggaggcccag gccttcattg agcgctactt tcagagcttc 2100
94  cccaaggtgc gggcctggat tgagaagacc ctggaggagg gcaggaggcg ggggtacgtg 2160
96  gagaccctct tcggccgccg ccgctacgtg ccagacctag aggcccgggt gaagagcgtg 2220
98  cgggaggcgg ccgagcgcat ggccttcaac atgcccgtcc agggcaccgc cgccgacctc 2280
100 atgaagctgg ctatggtgaa gctcttcccc aggctggagg aaatggggc caggatgctc 2340
102 cttcaggtcc acgacgagct ggtcctcgag gccccaaaag agagggcgga ggccgtggcc 2400
104 cggctggcca aggaggtcat ggaggggtg tatcccctgg ccgtgcccct ggaggtggag 2460
106 gtggggatag gggaggactg gctctccgcc aaggagtgat accacc                 2506
109 <210> SEQ ID NO: 2
110 <211> LENGTH: 2496
111 <212> TYPE: DNA
112 <213> ORGANISM: thermus flavus
114 <400> SEQUENCE: 2
115 atggcgatgc ttcccctctt tgagcccaaa ggccgcgtgc tcctggtgga cggccaccac 60
117 ctggcctacc gcaccttctt tgccctcaag ggcctcacca ccagccgcgg cgaacccgtt 120
119 caggcggtct acggcttcgc caaaagcctc ctcaaggccc tgaaggagga cggggacgtg 180
121 gtggtggtgg tctttgacgc caaggccccc tccttccgcc acgaggccta cgaggcctac 240
123 aaggcgggcc gggcccccac cccggaggac tttccccggc agctggccct catcaaggag 300
125 ttggtggacc tcctaggcct tgtgcggctg gaggttcccg gctttgaggc ggacgacgtg 360
127 ctggccaccc tggccaagcg ggcggaaaag gaggggtacg aggtgcgcat cctcactgcc 420
129 gaccgcgacc tctaccagct cctttcggag cgcatcgcca tcctccaccc tgaggggtac 480
131 ctgatcaccc cggcgtggct ttacgagaag tacggcctgc gcccggagca gtgggtggac 540
133 taccggggcc tggcggggga cccctcggat aacatccccg gggtgaaggg catcggggag 600
135 aagaccgccc agaggctcat ccgcgagtgg gggagcctgg aaaacctctt ccagcacctg 660
137 gaccaggtga gccctcctt gcgggagaag ctccaggcgg gcatggaggc cctggccctt 720
139 tcccggaagc tttcccaggt gcacactgac ctgccctgg aggtggactt cgggaggcgc 780
141 cgcacaccca acctggaggg tctgcgggct tttttggagc ggttggagtt tggaagcctc 840
143 ctccacgagt tcggcctcct ggaggggccg aaggcggcag aggaggcccc ctggcccct 900
145 ccggaagggg ctttttttggg cttttccttt tcccgtcccg agcccatgtg ggccgagctt 960
147 ctggcccctg ctggggcgtg ggaggggcgc ctccatcggg cacaagaccc ccttagggc 1020
149 ctgagggacc ttaagggggt gcggggaatc ctggccaagg acctggcggt tttggccctg 1080
151 cgggagggcc tggacctctt cccagaggac gaccccatgc tcctggccta ccttctggac 1140
153 ccctccaaca ccacccctga gggggtggcc cggcgttacg gggggagtg acggaggat 1200
155 gcggggggaga gggccctcct ggccgagcgc ctcttccaga ccctaaagga gcgccttaag 1260
157 ggagaagaac gcctgctttg gctttacgag gaggtggaga agccgctttc ccgggtgttg 1320
159 gcccggatgg aggccacggg ggtccggctg gacgtggcct acctccaggc cctctccctg 1380
161 gaggtggagg cggaggtgcg ccagctggag gaggaggtct tccgcctggc cggccacccc 1440
163 ttcaacctca actcccgcga ccagctggag cgggtgctct ttgacgagct gggcctgcct 1500
165 gccatcggca agacggagaa gacggggaaa cgctccacca gcgctgccgt gctggaggcc 1560
167 ctgcgagagg cccaccccat cgtggaccgc atcctgcagt accgggagct caccaagctc 1620
169 aagaacacct acatagaccc cctgcccgcc ctggtccacc ccaagaccgg ccggctccac 1680
171 acccgcttca accagacggc caccgccacg ggcaggcttt ccagctccga ccccaacctg 1740
173 cagaacatcc ccgtgcgcac ccctctgggc cagcgcatcc gccgagcctt cgtggccgag 1800
175 gagggctggg tgctggtggt cttggactac agccagattg agcttcgggt cctggcccac 1860
```

RAW SEQUENCE LISTING   DATE: 02/06/2005
PATENT APPLICATION: US/09/684,305   TIME: 12:32:05

          Input Set : A:\04447.app
          Output Set: N:\CRF4\02042005\I684305.raw

```
177 ctctccgggg acgagaacct gatccgggtc tttcaggagg ggagggacat ccacacccag 1920
179 accgccagct ggatgttcgg cgtttccccc gaaggggtag accctctgat gcgccgggcg 1980
181 gccaagacca tcaacttcgg ggtgctctac ggcatgtccg cccaccgcct ctccggggag 2040
183 ctttccatcc cctacgagga ggcggtggcc ttcattgagc gctacttcca gagctacccc 2100
185 aaggtgcggg cctggattga ggggaccctc gaggagggcc gccggcgggg gtatgtggag 2160
187 accctcttcg gccgccggcg ctatgtgccc gacctcaacg cccgggtgaa gagcgtgcgc 2220
189 gaggcggcgg agcgcatggc cttcaacatg ccggtccagg gcaccgccgc cgacctcatg 2280
191 aagctggcca tggtgcggct tttccccccgg cttcaggaac tggggcgag gatgcttttg 2340
193 caggtgcacg acgagctggt cctcgaggcc cccaaggacc gggcggagag ggtagccgct 2400
195 ttggccaagg aggtcatgga ggggggtctgg cccctgcagg tgcccctgga ggtggaggtg 2460
197 ggcctggggg aggactggct ctccgccaag gagtag             2496
200 <210> SEQ ID NO: 3
201 <211> LENGTH: 2504
202 <212> TYPE: DNA
203 <213> ORGANISM: Thermus thermophilus
205 <400> SEQUENCE: 3
206 atggaggcga tgcttccgct ctttgaaccc aaaggccggg tcctcctggt ggacggccac 60
208 cacctggcct accgcacctt cttcgccctg aagggcctca ccacgagccg gggcgaaccg 120
210 gtgcaggcgg tctacggctt cgccaagagc ctcctcaagg ccctgaagga ggacgggtac 180
212 aaggccgtct tcgtggtctt tgacgccaag gcccctcct tccgccacga ggcctacgag 240
214 gcctacaagg cggggagggc cccgaccccc gaggacttcc cccggcagct cgccctcatc 300
216 aaggagctgg tggacctcct ggggtttacc cgcctcgagg tccccggcta cgaggcggac 360
218 gacgttctcg ccaccctggc caagaaggcg gaaaaggagg ggtacgaggt cgcgcatcctc 420
220 accgccgacc gcgacctcta ccaactcgtc tccgaccgcg tcgccgtcct ccaccccgag 480
222 ggccacctca tcacccccgga gtggctttgg gagaagtacg gcctcaggcc ggagcagtgg 540
224 gtggacttcc gcgccctcgt gggggacccc tccgacaacc tccccgggggt caagggcatc 600
226 ggggagaaga ccgccctcaa gctcctcaag gagtggggaa gcctggaaaa cctcctcaag 660
228 aacctggacc gggtaaagcc agaaaacgtc cgggagaaga tcaaggccca cctggaagac 720
230 ctcaggctct ccttggagct ctcccgggtg cgcaccgacc tccccctgga ggtggacctc 780
232 gcccaggggc gggagcccga ccgggagggg cttagggcct tcctggagag ctggagttc 840
234 ggcagcctcc tccacgagtt cggcctcctg gaggcccccg cccccctgga ggaggccccc 900
236 tggccccccgc cggaagggggc cttcgtgggc ttcgtcctct cccgccccga gcccatgtgg 960
238 gcggagctta aagccctggc cgcctgcagg gacggccggg tgcaccgggc agcagacccc 1020
240 ttggcggggc taaaggacct caaggaggtc cggggcctcc tcgccaagga cctcgccgtc 1080
242 ttggcctcga gggaggggct agacctcgtg cccggggacg acccccatgct cctcgcctac 1140
244 ctcctggacc cctccaacac cacccccgag ggggtggcgc ggcgctacgg ggggggagtgg 1200
246 acggaggacg ccgcccaccg ggcctcctc tcggagaggc tccatcggaa cctccttaag 1260
248 cgcctcgagg gggaggagaa gctcctttgg ctctaccacg aggtggaaaa gcccctctcc 1320
250 cgggtcctgg cccacatgga ggccaccggg gtacggctgg acgtggccta ccttcaggcc 1380
252 cttttcctgg agcttgcgga ggagatccgc cgcctcgagg aggaggtctt ccgcttggcg 1440
254 ggccacccct tcaacctcaa ctcccgggac cagctggaaa gggtgctctt tgacgagctt 1500
256 aggcttcccg ccttggggaa gacgcaaaag acaggcaagc gctccaccag cgccgcggtg 1560
258 ctggaggccc tacgggaggc ccaccccatc gtggagaaga tcctccagca ccgggagctc 1620
260 accaagctca agaacaccta cgtggacccc ctcccaagcc tcgtccaccc gaggacgggc 1680
262 cgcctccaca cccgcttcaa ccagacggcc acggccacgg ggaggcttag tagctccgac 1740
264 cccaacctgc agaacatccc cgtccgcacc cccttgggcc agaggatccg ccgggccttc 1800
266 gtggccgagg cgggttgggc gttggtggcc ctggactata gccagataga gctccgcgtc 1860
268 ctcgcccacc tctccgggga cgaaaacctg atcagggtct tccaggaggg gaaggacatc 1920
270 cacacccaga ccgcaagctg gatgttcggc gtccccccgg aggccgtgga ccccctgatg 1980
```

```
272 cgccgggcgg ccaagacggt gaacttcggc gtcctctacg gcatgtccgc ccataggctc 2040
274 tcccaggagc ttgccatccc ctacgaggag gcggtggcct ttatagaggc tacttccaaa 2100
276 gcttccccaa ggtgcgggcc tggatagaaa agaccctgga ggaggggagg aagcggggct 2160
278 acgtggaaac cctcttcgga agaaggcgct acgtgcccga cctcaacgcc cgggtgaaga 2220
280 gcgtcaggga ggccgcggag cgcatggcct tcaacatgcc cgtccagggc accgccgccg 2280
282 acctcatgaa gctcgccatg gtgaagctct ccccccgcct ccgggagatg ggggcccgca 2340
284 tgctcctcca ggtccacgac gagctcctcc tggaggcccc ccaagcgcgg gccgaggagg 2400
286 tggcggcttt ggccaaggag gccatggaga aggcctatcc cctcgccgtg cccctggagg 2460
288 tggaggtggg gatgggggag gactggcttt ccgccaaggg ttag          2504
291 <210> SEQ ID NO: 4
292 <211> LENGTH: 832
293 <212> TYPE: PRT
294 <213> ORGANISM: Thermus aquaticus
296 <400> SEQUENCE: 4
297 Met Arg Gly Met Leu Pro Leu Phe Glu Pro Lys Gly Arg Val Leu Leu
298  1               5                   10                  15
300 Val Asp Gly His His Leu Ala Tyr Arg Thr Phe His Ala Leu Lys Gly
301              20                  25                  30
303 Leu Thr Thr Ser Arg Gly Glu Pro Val Gln Ala Val Tyr Gly Phe Ala
304          35                  40                  45
306 Lys Ser Leu Leu Lys Ala Leu Lys Glu Asp Gly Asp Ala Val Ile Val
307      50                  55                  60
309 Val Phe Asp Ala Lys Ala Pro Ser Phe Arg His Glu Ala Tyr Gly Gly
310  65                  70                  75                  80
312 Tyr Lys Ala Gly Arg Ala Pro Thr Pro Glu Asp Phe Pro Arg Gln Leu
313              85                  90                  95
315 Ala Leu Ile Lys Glu Leu Val Asp Leu Leu Gly Leu Ala Arg Leu Glu
316              100                 105                 110
318 Val Pro Gly Tyr Glu Ala Asp Asp Val Leu Ala Ser Leu Ala Lys Lys
319          115                 120                 125
321 Ala Glu Lys Glu Gly Tyr Glu Val Arg Ile Leu Thr Ala Asp Lys Asp
322      130                 135                 140
324 Leu Tyr Gln Leu Leu Ser Asp Arg Ile His Val Leu His Pro Glu Gly
325 145              150                 155                 160
327 Tyr Leu Ile Thr Pro Ala Trp Leu Trp Glu Lys Tyr Gly Leu Arg Pro
328              165                 170                 175
330 Asp Gln Trp Ala Asp Tyr Arg Ala Leu Thr Gly Asp Glu Ser Asp Asn
331              180                 185                 190
333 Leu Pro Gly Val Lys Gly Ile Gly Glu Lys Thr Ala Arg Lys Leu Leu
334          195                 200                 205
336 Glu Glu Trp Gly Ser Leu Glu Ala Leu Leu Lys Asn Leu Asp Arg Leu
337      210                 215                 220
339 Lys Pro Ala Ile Arg Glu Lys Ile Leu Ala His Met Asp Asp Leu Lys
340 225              230                 235                 240
342 Leu Ser Trp Asp Leu Ala Lys Val Arg Thr Asp Leu Pro Leu Glu Val
343              245                 250                 255
345 Asp Phe Ala Lys Arg Arg Glu Pro Asp Arg Glu Arg Leu Arg Ala Phe
346              260                 265                 270
348 Leu Glu Arg Leu Glu Phe Gly Ser Leu Leu His Glu Phe Gly Leu Leu
```

```
349           275              280              285
351 Glu Ser Pro Lys Ala Leu Glu Glu Ala Pro Trp Pro Pro Pro Glu Gly
352     290              295              300
354 Ala Phe Val Gly Phe Val Leu Ser Arg Lys Glu Pro Met Trp Ala Asp
355 305              310              315              320
357 Leu Leu Ala Leu Ala Ala Ala Arg Gly Gly Arg Val His Arg Ala Pro
358              325              330              335
360 Glu Pro Tyr Lys Ala Leu Arg Asp Leu Lys Glu Ala Arg Gly Leu Leu
361              340              345              350
363 Ala Lys Asp Leu Ser Val Leu Ala Leu Arg Glu Gly Leu Gly Leu Pro
364              355              360              365
366 Pro Gly Asp Asp Pro Met Leu Leu Ala Tyr Leu Leu Asp Pro Ser Asn
367     370              375              380
369 Thr Thr Pro Glu Gly Val Ala Arg Arg Tyr Gly Gly Glu Trp Thr Glu
370 385              390              395              400
372 Glu Ala Gly Glu Arg Ala Ala Leu Ser Glu Arg Leu Phe Ala Asn Leu
373              405              410              415
375 Trp Gly Arg Leu Glu Gly Glu Glu Arg Leu Leu Trp Leu Tyr Arg Glu
376              420              425              430
378 Val Glu Arg Pro Leu Ser Ala Val Leu Ala His Met Glu Ala Thr Gly
379              435              440              445
381 Val Arg Leu Asp Val Ala Tyr Leu Arg Ala Leu Ser Leu Glu Val Ala
382     450              455              460
384 Glu Glu Ile Ala Arg Leu Glu Ala Glu Val Phe Arg Leu Ala Gly His
385 465              470              475              480
387 Pro Phe Asn Leu Asn Ser Arg Asp Gln Leu Glu Arg Val Leu Phe Asp
388              485              490              495
390 Glu Leu Gly Leu Pro Ala Ile Gly Lys Thr Glu Lys Thr Gly Lys Arg
391              500              505              510
393 Ser Thr Ser Ala Ala Val Leu Glu Ala Leu Arg Glu Ala His Pro Ile
394              515              520              525
396 Val Glu Lys Ile Leu Gln Tyr Arg Glu Leu Thr Lys Leu Lys Ser Thr
397     530              535              540
399 Tyr Ile Asp Pro Leu Pro Asp Leu Ile His Pro Arg Thr Gly Arg Leu
400 545              550              555              560
402 His Thr Arg Phe Asn Gln Thr Ala Thr Ala Thr Gly Arg Leu Ser Ser
403              565              570              575
405 Ser Asp Pro Asn Leu Gln Asn Ile Pro Val Arg Thr Pro Leu Gly Gln
406              580              585              590
408 Arg Ile Arg Arg Ala Phe Ile Ala Glu Glu Gly Trp Leu Leu Val Ala
409              595              600              605
411 Leu Asp Tyr Ser Gln Ile Glu Leu Arg Val Leu Ala His Leu Ser Gly
412     610              615              620
414 Asp Glu Asn Leu Ile Arg Val Phe Gln Glu Gly Arg Asp Ile His Thr
415 625              630              635              640
417 Glu Thr Ala Ser Trp Met Phe Gly Val Pro Arg Glu Ala Val Asp Pro
418              645              650              655
420 Leu Met Arg Arg Ala Ala Lys Thr Ile Asn Phe Gly Val Leu Tyr Gly
421     660              665              670
```

RAW SEQUENCE LISTING ERROR SUMMARY   DATE: 02/06/2005
PATENT APPLICATION: US/09/684,305   TIME: 12:32:06

Input Set : A:\04447.app
Output Set: N:\CRF4\02042005\I684305.raw

Please Note:
**Use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.**

Seq#:7; N Pos. 4,5,181,182,190,366,617,628,685,714,722,738,784,1022,1029
Seq#:7; N Pos. 1038,1053,1098,1105,1206,1227,1244,1251,1252,1253,1350,1380
Seq#:7; N Pos. 1497,1530,1569,1572,1641,1653,1655,1770,1812,2319,2346,2396
Seq#:8; Xaa Pos. 2,63,109,186,205,209,227,228,233,240,243,244,247,260,290
Seq#:8; Xaa Pos. 329,336,340,368,414,417,418,431,551,605,773,794,798,823
Seq#:8; Xaa Pos. 833

VERIFICATION SUMMARY   DATE: 02/06/2005
PATENT APPLICATION: **US/09/684,305**   TIME: 12:32:06

Input Set : **A:\04447.app**
Output Set: **N:\CRF4\02042005\I684305.raw**

L:970 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7 after pos.:0
M:341 Repeated in SeqNo=7
L:1199 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:8 after pos.:0
M:341 Repeated in SeqNo=8